DOCUMENT RESUME

ED 192 262 CS 005 633

AUTHOR Sakiey, Elizabeth: And Others
TITLE A Syllable Frequency Ccunt.
PUB DATE [77]
NOTE 27p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computational Linguistics: *Language Research: Reading Instruction: *Syllables: *Word Frequency: *Written Language
IDENTIFIERS *Frequency Data

ABSTRACT

A frequency count of the syllables in an edited list of the 5,000 most common English words generated an unweighted list of the most common 322 syllables in the words and a list of 290 syllables weighted by the frequency of occurrence of words of which the syllables were constituents. The unweighted list ccntained 222 nonword syllables and 100 word syllables. The weighted list contained 190 word syllables and 100 nonword syllables. The latter 290 syllables account for 72% of the approximately six million syllable tokens in the 5,000 most common English words. The lists are a potential data base for the development of curriculum materials in reading, spelling, and other areas. They are also a potential data base for investigations and technologies in readability, computer translation, verbal learning, and language acquisition. (The lists are appended.) (Author/FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192262

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A SYLLABLE FREQUENCY COUNT

by

Elizabeth Sakiey

Reading/Speech Correction Department

Glassboro State College

Glassboro, NJ 08028

and

Edward B. Fry

Reading Center

Graduate School of Education

Rutgers University

New Brunswick, NJ 08903

and

Albert Goss

Psychology Department

Douglass College

Rutgers University

New Brunswick, NJ 08903

and

Barry Loigman

Psychology Department

Glassboro State College

Glassboro, NJ 08028

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Elizabeth Sakiey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS005633

A SYLLABLE FREQUENCY COUNT

ABSTRACT

A frequency count of the syllables in an edited list of the 5,000 most common English words generated an unweighted list of the most common 322 syllables in 5,000 words and a list of 290 syllables weighted by the frequency of occurrence of words of which the syllables were constituents. The unweighted list contained 222 nonword-syllables and 100 word-syllables. The weighted list contained 190 word-syllables and 100 nonword-syllables. The latter 290 syllables account for 72 percent of the 5,890,868 syllable tokens in the 5,000 most common English words. These lists are seen as potential base data for the development of curriculum materials in reading, spelling, and other areas. They are also seen as potential base data for investigations and technologies in readability, computer translation, verbal learning, and language acquisition.

A SYLLABLE FREQUENCY COUNT

This study updates and refines certain aspects of basic knowledge about the syllable. Historically, the syllable was one of the mainstays of the teaching of reading; it played a prominent role in the methodology of the 18th century New England Primer. At various times subsequently in the history of instruction the syllable has been featured in reading, spelling, and writing instruction. Syllables continue to be of minor importance in those areas. They are more important in dictionary use, speech correction, typography, computer translation, readability, verbal learning, and language research.

The problem of this study was to develop a syllable rank-order list. It attempts to answer the questions, "What is the commonest syllable? What is the next commonest syllable?" etc.

These questions pose another question, "What does common mean?" In terms of this study, it means both an unweighted rank-order list and a weighted rank-order list. The unweighted list presents the number of occurrences of each different syllable in a list of the 5,000 most common words in an edited list based on the Carroll, Davies, and Richman or American Heritage (1971) count. The weighted list reflects the frequency of occurrence of each syllable in the 5,000 words weighted by the number of times each of those words appeared in 5,088,721 running words. These two lists are then compared.

Background

The last major syllable count was done by Osburn (1954). He listed the 15 commonest initial, medial, and final syllables. Unfortunately, the bulk of his syllable count on the 9,000 polysyllabic words of the now dated Rinsland (1945) list did not find its way into the professional literature and, for most practical purposes, is lost.

Three decades earlier Dewey counted phonic syllables (spoken syllables) in 100,000 printed words of adult material. He ranked the 220 most common of the 4,400 different syllables obtained (1923). Later he appended initial, medial, and final positions for these syllables (1950).

Dolch (1938) obtained 8,509 nonword syllables in 3,931 polysyllabic words found in 14,000 running words of elementary texts. There were 1,255 different syllables. His published alphabetical list of the 100 commonest syllables does not answer the question of which is the commonest syllable. The list of 100 syllables is all that is apparently available (Dolch, 1940). Additionally, the Dolch count, like the Dewey and Osburn counts, was an unweighted one.

Data Base

The Carroll et. al. frequencies that served as the data base were generated by a computerized count of occurrences of different words in 5,088,721 running words. These words were in 10,043 five-hundred word samples of reading materials used in grades three through nine in 22 different categories. This count yielded 86,741 different words or

types. A type might not be a word because numerals such as "1905" and initializations such as "USA" are types. Each occurrence of a type is a token.

A token is the frequency of a word. For example, the type "the" occurred 373,123 times, or "the" had 373,123 tokens.

Method

The 86,741 different words (types) of the Carroll et. al. count, together with their frequencies of occurrence, were on computer data tape deposited at the National Archives for Linguistics in Arlington, Virginia.[1] Punched cards were generated from these tapes for 52,000 words that occurred more than once per five million.

The first phase was to edit the Carroll et. al. list. They defined a word as a group of symbols with a space on either side. This facilitated selection, scoring, and counting by the computer. It also had the advantage of nonjudgmental objectivity. However, for the purposes of this study it had the obvious drawback of counting nonwords such as "&," "$100," "1945," and "USA" as words. Also, words which differed only in case of letters such as "Run," "run," and "RUN," were counted as three different words with three different frequencies of occurrence for each word.

Hand editing was necessary to eliminate nonalphabetic symbols and nonwords. Hand editing was also used to combine frequencies for different graphic forms of the same word, such as "Run" and "run," into one word. However, inflected forms of the word, such as "runs," "run-

ning," and "ran" were not combined; this would have omitted syllables such as "ed" and "ing."

After hand editing, 44,174 word cards remained. Of these words, the count was on the 5,000 most frequent. These words account for 4,513,777 occurrences or 89 percent of the word tokens in the 5,088,721 running words. These 5,000 high frequency words are all the words that occur more than approximately 15 times per million running words.

The 5,000 most frequent words were then divided into graphemic syllables. Graphemic syllables refer to syllables whose boundaries were determined for graphic (written) use. The American Heritage School Dictionary (1977) was the authority for this syllabification.

Phonetic syllables are those syllables used in the phonetic or pronunciation part of a dictionary definition; they are said to be more akin to speech patterns. There is a lack of agreement among dictionaries on syllable boundaries, particularly for phonetic syllables.

The computer programs used were based on Fortran Sort-Merge and the Statistical Package from the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, and Brent, 1975).

Results

Unweighted List

There were 9,358 graphemic syllables or syllable tokens in the 5,000 words, an average of about 1.9 syllables per word. These 9,358 syllables were comprised of 3,402 different syllables or syllable types.

The different syllables may be divided into word-syllables and nonword syllables. A word-syllable is one like "the" which is a simple word and may also be part of a larger word like "theater." A nonword-syllable is a syllable like "ing," which never occurs as a single word. Of the 3,402 different syllables, 1,982 or 58 percent are word-syllables and 1,420 or 42 percent are nonword-syllables.

The 322 most frequent graphemic syllables are presented in Table 1. These are syllables that occurred in five or more of the 5,000 words. Of these syllables, 222 or 69 percent are nonword-syllables and 100 or 31 percent are word-syllables. The five most frequent syllables, "ing," "er," "a," "ly," and "ed," are all nonword-syllables. The first word syllable, "in," has a rank of 10; the second word-syllable, "an," has a rank of 22.

The sum of the frequencies of the 222 nonword-syllables is 3,329 or 36 percent of the 9,358 syllable tokens in the 5,000 words. The sum of the frequencies of the 100 word-syllables is 1,496 or 16 percent of the syllable tokens. Thus, 322 syllables account for 52 percent of the 9,358 syllable tokens.

Weighted List

The rank, weighted frequency, and unweighted frequency of each of the 290 syllables with a weighted frequency of 3,000 or more are in Table 2. Of these 290 syllables, 190 or 66 percent are word-syllables and 100 or 34 percent are nonword-syllables. The most frequent syllable of the unweighted list, "ing," drops to rank seven in the weighted list, and "er" drops from rank two to rank eight. Of the first 100 syllables, 72

are word-syllables.

When 3,402 syllable types in the 5,000 words are weighted for frequency of the words in which they appear, there are 5,890,868 syllable tokens. The common 190 word-syllables in Table 2 account for 3,260,280 or 55 percent of these tokens. The 100 nonword-syllables account for 977,456 or 17 percent of these tokens. All 290 syllables account for 4,237,736 or 72 percent of the 5,890,868 tokens.

Discussion

The importance of the data presented here is that a relatively small number of syllables account for a large proportion of our written language.

In the weighted list, a mere 290 word- and nonword-syllables account for 72 percent of the 5,890,896 syllable tokens in the 5,000 most common words weighted for their frequency. This reflects the J-curve distribution of the weighted frequencies in which a few types occur with high frequency and many types occur with moderate to low frequencies.

High frequency whole words continue to be important. Of the 290 weighted syllables, 190 are word-syllables which account for 55 percent of the 5,890,896 tokens or more than half of all written English. This supports the efficacy of teaching a high frequency or basic sight vocabulary such as those developed by Dolch (1936) and Fry (1957) as both the Dolch and Fry list contain a high percentage of the same one-syllable words. Furthermore, the 190 word-syllables presented in Table 2 make up about half of all written English since the 190 words make up 49% of the

five million running words in the Carroll count.

Another finding of this study is that just 100 non-word syllables of the weighted list account for an additional 17 percent of the 5,890,896 syllable tokens. Thus, the 100 syllables account for approximately another sixth of all written English.

These findings cannot be compared with those of earlier counts. Dolch and Osburn counted only polysyllabic words; Dewey used phonemic syllables, and all these were unweighted counts.

The unweighted count presented here provides a different orientation than the weighted count. When one looks at the unweighted list, in this analysis of 5,000 common words, the 222 nonword-syllables of Table 1 make up 36 percent of the total syllables, and 100 word-syllables make up 16 percent of the total. Thus, nonword-syllables assume greater importance on the unweighted list than on the weighted list.

Other researchers provide evidence that the syllable is a viable unit for at least some phases of literacy instruction. Savin and Bever (1970) demonstrated "that even for literate adults the syllable is a far more natural, more easily available perceptual unit than is the phoneme" (p. 322). Using both urban and suburban kindergarten-age children, Allen, Rozin, and Gleitman (1972) found that prereaders were more likely to recognize a meaningful word when segmented syllabically than when segmented phonemically.

Blending of two syllables was found to be easier than blending two

phonemes (Brown, 1971) for children of ages 56 to 80 months. Children 4 to 6 years old had more difficulty analyzing words into phonemes than syllables (Liberman, Shakweiler, Carter, & Fischer, 1972). Finally, tachistoscopic studies have shown that the syllable functions as a single perceptual unit (Spoehr & Smith, 1973).

Ruddell (1976) called for "A study of various decoding units ... and the relationship between these units and early reading success" (p. 35). This is only one of the areas that high-frequency syllable lists might be employed. They might be used in the development of curriculum materials for developmental and remedial reading as well as in spelling and typing instruction. Syllable frequency lists may also be useful in a wide variety of investigation of and technologies of readability, spelling reform, verbal learning, language acquisition, and computer translation.

The unique contribution of this study was to present a weighted syllable frequency count and the finding that a relatively small number of syllables make up a fairly substantial portion of all written language.

Footnote

1. The authors would like to thank A. Hood Roberts and Nancy Jokovich at the Center for Applied Linguistics for helping in the acquisition of the American Heritage computer tapes from their archival holdings. The authors also wish to thank Anka Wagner for her aid in computer programming. The authors are greatly indebted to John Carroll, Peter Davies, and Barry Richman, along with the American Heritage Publishing Company and Houghton-Mifflin for setting a new, high standard in word counts and for making their data available.

TABLE I. THE 322 MOST COMMON UNWEIGHTED GRAPHEMIC SYLLABLES IN THE ENGLISH LANGUAGE RANKED IN ORDER OF FREQUENCY IN THE 5,000 MOST FREQUENT WORDS

Rank	Syl.	F.	Rank	Syl.	F.	Rank	Syl.	F.	Rank	Syl.	F.
1.	ing	230		ty	39	45.	at	21	67.	as	15
2.	er	129	24.	ry	37		dis	21		col	15
3.	a	124	25.	u	36	47.	ca	20		fi	15
4.	ly	119	26.	ti	35		cal	20		ful	15
5.	ed	114	27.	ri	32		man	20		ger	15
6.	i	112	28.	be	30	50.	ap	19		low	15
7.	es	98	29.	per	29		po	19		ni	15
8.	re	90		to	29		sion	19		par	15
9.	tion	83	31.	pro	28		vi	19		son	15
10.	in	78	32.	ac	26	54.	el	18		tle	15
11.	e	67		ad	26		est	18	77.	day	14
12.	con	64		ar	26		la	18		ny	14
13.	y	63		ers	26		lar	18		pen	14
14.	ter	60		ment	26		pa	18		pre	14
15.	ex	58		or	26		ture	18		tive	14
16.	al	56		tions	26	60.	for	17	82.	car	13
17.	de	55	39.	ble	25		is	17		ci	13
18.	com	51		der	25		mer	17		mo	13
	o	51	41.	ma	23		pe	17		on	13
20.	di	46	42.	na	22		ra	17		ous	13
21.	en	42		si	22	65.	so	16		pi	13
22.	an	39		un	22		ta	16		se	13

(Continued)

TABLE I (Continued)

Rank	Syl.	F.	Rank	Syl.	F.	Rank	Syl.	F.	Rank	Syl.	F.
	ten	13		ings	10		sen	9		mar	8
	to:	13		mag	10		side	9		mis	8
	ver	13		ments	10		tal	9		my	8
92.	ber	12		set	10		tic	9		nal	8
	can	12		some	10		ties	9		ness	8
	dy	12		sub	10		ward	9		ning	8
	et	12		sur	10	141.	age	8		n't	8
	it	12		ters	10		ba	8		nu	8
	mu	12		tu	10		but	8		oc	8
	no	12	121.	af	9		cit	8		pres	8
	ple	12		au	9		cle	8		sup	8
100.	cu	11		cy	9		co	8		te	8
	fac	11		fa	9		cov	8		ted	8
	fer	11		im	9		da	8		tem	8
	gen	11		li	9		dif	8		tin	8
	ic	11		lo	9		ence	8		tri	8
	land	11		men	9		ern	8		tro	8
	light	11		min	9		eve	8		up	8
	ob	11		mon	9		hap	8		va	8
	of	11		op	9		ies	8		ven	8
	pos	11		out	9		ket	8		vis	8
	tain	11		rec	9		lec	8		am	8
111.	den	10		ro	9		main	8		bor	8

(Continued)

TABLE I (Continued)

Rank	Syl.	F.	Rank	Syl.	F.	Rank	Syl.	F.	Rank	Syl.	F.
	by	8		way	7		let	6		vid	6
	cat	8	205.	ate	6		long	6		wil	6
183.	cent	7		bet	6		mat	6		win	6
	ev	7		bles	6		meas	6		won	6
	gan	7		bod	6		mem	6		work	6
	gle	7		cap	6		mul	6	255.	act	5
	head	7		cial	6		ner	6		ag	5
	high	7		cir	6		play	6		air	5
	il	7		cor	6		ples	6		als	5
	lu	7		coun	6		ply	6		bat	5
	me	7		cus	6		port	6		bi	5
	nor	7		dan	6		press	6		cate	5
	part	7		dle	6		sat	6		cen	5
	por	7		ef	6		sec	6		char	5
	read	7		end	6		ser	6		come	5
	rep	7		ent	6		south	6		cul	5
	su	7		ered	6		sun	6		ders	5
	tend	7		fin	6		the	6		east	5
	ther	7		form	6		ting	6		fect	5
	ton	7		go	6		tra	6		fish	5
	try	7		har	6		tures	6		fix	5
	um	7		ish	6		val	6		gi	5
	ure	7		lands	6		var	6		grand	5

(Continued)

TABLE I (Continued)

Rank	Syl.	F.	Rank	Syl.	F.	Rank	Syl.	F.	Rank	Syl.	F.
	great	5		mal	5		ried	5		sug	5
	heav	5		mi	5		round	5		tel	5
	ho	5		mil	5		row	5		tom	5
	hunt	5		moth	5		sa	5		tors	5
	ion	5		near	5		sand	5		tract	5
	its	5		nel	5		self	5		tray	5
	jo	5		net	5		sent	5		us	5
	lat	5		new	5		ship	5		vel	5
	lead	5		one	5		sim	5		west	5
	lect	5		point	5		sions	5		where	5
	lent	5		prac	5		sis	5		writ	5
	less	5		ral	5		sons	5			
	lin	5		rect	5		stand	5			

TABLE II. THE 290 MOST COMMON SYLLABLES IN THE ENGLISH LANGUAGE RANKED IN ORDER OF FREQUENCY IN 5,088,721 RUNNING WORDS

		Frequencies				Frequencies	
Rank	Syl.	Weighted[b]	Unwtd.[a]	Rank	Syl.	Weighted[b]	Unwtd.[a]
1.	the	374,747	6	23.	are	35,502	1
2.	a	191,870	124	24.	with	33,817	3
3.	of	150,386	11	25.	ter	32,154	60
4.	to	143,284	29	26.	his	30,244	3
5.	and	133,921	1	27.	at	27,710	21
6.	in	131,121	78	28.	or	27,697	26
7.	ing	71,958	230	29.	they	27,627	1
8.	er	64,846	129	30.	al	27,231	56
9.	is	64,816	17	31.	ed	26,367	114
10.	i	59,077	112	32.	es	24,969	98
11.	be	57,052	30	33.	this	23,316	1
12.	you	50,999	1	34.	from	22,810	1
13.	it	50,254	12	35.	one	22,644	5
14.	that	47,462	1	36.	have	22,444	2
15.	y	46,700	63	37.	e	21,956	67
16.	on	46,634	13	38.	by	21,746	7
17.	he	46,532	3	39.	man	21,081	20
18.	for	44,258	17	40.	tion	20,975	83
19.	was	41,685	2	41.	had	20,802	2
20.	ly	40,685	119	42.	not	20,090	3
21.	an	40,572	39	43.	but	19,993	3
22.	as	35,619	15	44.	can	19,683	12

(Continued)

TABLE II (Continued)

		Frequencies				Frequencies	
Rank	Syl.	Weighted[b]	Unwtd.[a]	Rank	Syl.	Weighted[b]	Unwtd.[a]
45.	so	19,242	16	67.	which	14,016	1
46.	re	18,799	90	68.	do	13,744	3
47.	some	18,558	10	69.	she	13,657	1
48.	what	18,301	3	70.	their	13,258	1
49.	o	18,094	51	71.	them	12,959	2
50.	were	17,207	2	72.	if	12,912	1
51.	oth	17,172	4	73.	will	12,873	2
52.	all	17,042	1	74.	di	12,810	46
53.	out	16,892	9	75.	him	12,542	2
54.	we	16,474	3	76.	bout	12,507	1
55.	ry	16,278	37	77.	com	12,439	51
56.	your	16,214	2	78.	pie	12,420	12
57.	when	16,183	2	79.	u	12,248	36
58.	there	15,692	2	80.	then	12,026	1
59.	how	15,422	3	81.	her	11,814	2
60.	said	15,309	1	82.	no	11,763	12
61.	up	15,291	8	83.	words	11,707	1
62.	de	14,977	55	84.	these	11,611	1
63.	ver	14,953	13	85.	con	11,598	64
64.	ex	14,710	58	86.	way	11,406	7
65.	each	14,290	1	87.	per	11,315	29
66.	en	14,158	42	88.	would	11,191	1

(Continued)

TABLE II (Continued)

Rank	Syl.	Frequencies		Rank	Syl.	Frequencies	
		Weighted[b]	Unwtd.[a]			Weighted[b]	Unwtd.[a]
89.	low	10,801	15	111.	peo	8,281	3
90.	un	10,748	22	112.	its	8,197	5
91.	like	10,664	4	113.	ble	8,159	25
92.	long	10,616	6	114.	than	8,057	2
93.	has	10,469	2	115.	af	7,934	9
94.	two	10,144	1	116.	ers	7,911	26
95.	my	10,142	8	117.	may	7,836	3
96.	more	10,130	2	118.	word	7,804	1
97.	go	10,055	6	119.	first	7,776	1
98.	write	9,974	2	120.	mer	7,726	17
99.	der	9,844	25	121.	wa	7,699	3
100.	tle	9,636	15	122.	ten	7,661	13
101.	could	9,464	2	123.	been	7,651	1
102.	ber	9,397	12	124.	who	7,608	1
103.	did	9,276	3	125.	ment	7,535	26
104.	ty	9,080	39	126.	use	7,529	3
105.	see	8,981	3	127.	now	7,465	1
106.	num	8,955	4	128.	ti	7,451	35
107.	day	8,949	14	129.	pro	7,447	28
108.	time	8,643	3	130.	down	7,419	3
109.	most	8,372	3	131.	find	7,313	2
110.	make	8,340	1	132.	ar	7,285	26

(Continued)

TABLE II (Continued)

		Frequencies				Frequencies	
Rank	Syl.	Weighted[b]	Unwtd.[a]	Rank	Syl.	Weighted[b]	Unwtd.[a]
133.	me	7,256	7	155.	pa	6,105	18
134.	ma	7,231	23	156.	let	5,998	6
135.	new	7,190	5	157.	tions	5,981	26
136.	lit	7,157	4	158.	just	5,939	2
	made	7,157	2	159.	work	5,932	6
138.	get	7,029	4	160.	know	5,926	2
139.	ri	6,956	32	161.	our	5,922	2
140.	thing	6,817	4	162.	ther	5,913	7
141.	eve	6,806	8	163.	through	5,909	2
142.	us	6,608	5	164.	try	5,898	7
143.	sen	6,601	9	165.	fore	5,870	3
144.	read	6,584	7	166.	called	5,789	1
145.	come	6,467	5	167.	great	5,737	5
146.	came	6,418	2	168.	est	5,728	18
147.	where	6,413	5	169.	fa	5,705	9
148.	ture	6,411	18	170.	good	5,629	3
149.	look	6,286	2	171.	used	5,611	1
150.	back	6,252	4	172.	la	5,608	18
151.	side	6,228	9	173.	land	5,597	11
152.	fer	6,211	11	174.	part	5,551	7
153.	dif	6,201	8	175.	car	5,491	13
154.	round	6,168	5	176.	el	5,474	18

(Continued)

TABLE II (Continued)

		Frequencies				Frequencies	
Rank	Syl.	Weighted[b]	Unwtd.[a]	Rank	Syl.	Weighted[b]	Unwtd.[a]
177.	think	5,445	2	199.	tain	4,716	11
178.	n't	5,433	8	200.	ning	4,715	8
179.	much	5,388	1	201.	cause	4,714	2
180.	si	5,218	22	202.	ways	4,685	3
181.	set	5,217	10	203.	col	4,627	15
182.	ent	5,187	6	204.	am	4,621	7
183.	ven	5,098	8		par	4,621	15
184.	ev	5,096	7	206.	dis	4,549	21
185.	too	5,074	1	207.	small	4,509	3
186.	men	5,064	9	208.	air	4,492	5
187.	old	5,034	3	209.	three	4,479	1
188.	name	5,024	1	210.	put	4,455	4
189.	ac	5,012	26	211.	say	4,442	2
190.	ca	5,008	20	212.	ern	4,409	8
191.	does	5,001	2	213.	help	4,387	3
192.	sound	4,964	2	214.	self	4,380	5
193.	fol	4,932	4	215.	ny	4,372	14
194.	right	4,931	2	216.	times	4,361	2
195.	place	4,883	4	217.	well	4,350	2
196.	ful	4,867	15	218.	cit	4,331	8
197.	son	4,747	15	219.	must	4,307	1
198.	na	4,726	22	220.	want	4,292	2

(Continued)

TABLE II (Continued)

		Frequencies				Frequencies	
Rank	Syl.	Weighted[b]	Unwtd.[a]	Rank	Syl.	Weighted[b]	Unwtd.[a]
221.	big	4,254	3	243.	high	3,861	7
222.	take	4,252	2	244.	light	3,809	11
223.	po	4,247	19	245.	head	3,795	7
224.	such	4,223	1	246.	coun	3,770	6
225.	cal	4,197	20	247.	mon	3,712	9
226.	here	4,192	1	248.	pe	3,698	17
227.	why	4,158	1	249.	near	3,677	5
228.	tell	4,137	2	250.	lar	3,654	18
229.	went	4,132	1	251.	por	3,595	7
230.	line	4,131	4	252.	fi	3,579	15
231.	pen	4,121	14	253.	bers	3,563	3
232.	mu	4,097	12	254.	sec	3,557	6
233.	things	4,078	1	255.	ap	3,545	19
234.	moth	4,077	5	256.	stud	3,491	4
235.	gain	4,065	2	257.	found	3,477	2
236.	end	4,027	6	258.	should	3,470	1
237.	pic	4,017	4	259.	ad	3,431	21
238.	im	4,001	9	260.	still	3,424	1
239.	to	3,989	16	261.	form	3,414	6
240.	years	3,975	1	262.	need	3,412	2
241.	off	3,875	1	263.	play	3,392	6
242.	name	3,864	2	264.	world	3,383	1

(Continued)

TABLE II (Continued)

		Frequencies				Frequencies	
Rank	Syl.	Weighted[b]	Unwtd.[a]	Rank	Syl.	Weighted[b]	Unwtd.[a]
265.	home	3,370	1	278.	mean	3,201	4
266.	give	3,369	1	279.	looked	3,197	1
267.	tween	3,344	1	280.	earth	3,177	2
268.	own	3,313	3	281.	ters	3,174	10
269.	gan	3,285	7	282.	cov	3,165	8
270.	bod	3,276	6	283.	ger	3,147	15
271.	add	3,254	3	284.	last	3,132	2
272.	tence	3,252	2	285.	nit	3,095	3
273.	ward	3,250	9	286.	show	3,089	2
274.	hap	3,238	8	287.	might	3,077	2
	nev	3,238	2	288.	stand	3,058	5
276.	ure	3,208	7	289.	house	3,054	2
277.	mem	3,206	6	290.	got	3,051	3

TABLE I NOTE: This table contains all syllables that occurred five or more times in the 5,000 highest frequency English words. These 322 syllables comprise 52 percent of all the syllables (tokens) in the 5,000 words.

TABLE II NOTE:

a. Unweighted Frequency means number of occurrences in the 5,000 different words. This is the same as Table I count.

b. Weighted Frequency means number of occurrences per five million running words using words from the 5,000 word list. For example, the 189th syllable "ac" appeared in 26 different words of the 5,000 word list and these 26 words were used a total of 5,012 times in five million running words.

These 290 syllables account for 72 percent of all the syllables (tokens) in the 5,000 most common words.

REFERENCES

Allen, M.W., Rozin, P., & Gleitman, L.R. A test of the blending abilities of kindergarten children using syllable and phoneme segments. Unpublished manuscript, University of Pennsylvania, 1972. Cited by L.R. Gleitman & P. Rozin, Teaching reading by use of a syllabary. Reading Research Quarterly, Summer 1973, 8, 447-483.

The American Heritage School Dictionary. Boston: Houghton-Mifflin, 1977.

Brown, D.L. Some linguistic dimensions in auditory blending. In G. Green (Ed.), Reading: The right to participate. Yearbook of the National Reading Conference, 1971, 20, 227-236. Cited by L.R. Gleitman & P. Rozin, Teaching reading by use of a syllabary, Reading Research Quarterly, Summer 1973, 8, 447-483.

Carroll, J.B., Davies, P., & Richman, B. American heritage word frequency book. Boston: Houghton-Mifflin, 1971.

Dewey, G. Relativ frequency of English speech sounds. Cambridge: Harvard University Press, 1923. (Reprinted in 1950.)

Dolch, E.W. A Basic Sight Vocabulary, Elementary School Journal, 1936, 36, 456-460; 37, 268-272.

Dolch, E.W. Phonics and polysyllables. Elementary English Review, 1938, 15, 120-124.

Dolch, E.W. Sight Syllables versus letter phonics. Elementary School Journal, 1940, 41, 38-42.

Fry, Edward. "Developing a Word List for Remedial Reading." Elementary English, November, 1957.

Liberman, I.Y., Shankweiler, D., Carter, B., & Fischer, W.F. Reading and the awareness of linguistic segments. Unpublished manuscript, University of Connecticut, 1972. Cited by L.R. Gleitman & P. Rozin, Teaching reading by use of a syllabary. Reading Research Quarterly, Summer, 1973, 8, 447-483.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. Statistical package for the social sciences. (2nd ed.) New York: McGraw-Hill, 1975.

Osburn, W.J. Teaching spelling by teaching syllables and root words. Elementary School Journal, 1954, 55, 32-41.

Rinsland, H.D. A basic vocabulary of elementary school children. New York: Macmillan, 1945.

Ruddell, R.B. Language acquisition and the reading process. In H. Singer & R.B. Ruddell (Eds.) Theoretical models and processes of reading. Newark, Del.: International Reading Association, 1976, pp. 22-38.

Savin, H.B., & Bever, T.G. The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 295-302.

Spoehr, K.T., & Smith, E.E. The role of syllables in perceptual processing. Cognitive Psychology, 1973, 5, 71-89.